

<https://helda.helsinki.fi>

The effect of reducing EEG electrode number on the visual interpretation of the human expert for neonatal seizure detection

Stevenson, Nathan J.

2018-01

Stevenson , N J , Lauronen , L & Vanhatalo , S 2018 , ' The effect of reducing EEG electrode number on the visual interpretation of the human expert for neonatal seizure detection ' , Clinical Neurophysiology , vol. 129 , no. 1 , pp. 265-270 . <https://doi.org/10.1016/j.clinph.2017.10.031>

<http://hdl.handle.net/10138/298237>

<https://doi.org/10.1016/j.clinph.2017.10.031>

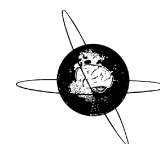
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.



The effect of reducing EEG electrode number on the visual interpretation of the human expert for neonatal seizure detection



Nathan J. Stevenson*, Leena Lauronen, Sampsa Vanhatalo

Clinicum, Faculty of Medicine, University of Helsinki, Finland

BABA Center and Department of Children's Clinical Neurophysiology, HUS Medical Imaging Center, Helsinki University Central Hospital, Finland

ARTICLE INFO

Article history:

Accepted 19 October 2017

Available online 13 November 2017

Keywords:

Neonatal EEG

Inter-observer agreement

Reliability

Seizure detection

Brain monitoring

Amplitude integrated EEG

HIGHLIGHTS

- Expert visual interpretation of multichannel EEG is standard for neonatal seizure detection.
- A reduced number of electrodes is often used in neonatal intensive care monitoring.
- Reducing the number of electrodes results in a significant reduction in seizure detection.

ABSTRACT

Objectives: To measure changes in the visual interpretation of the EEG by the human expert for neonatal seizure detection when reducing the number of recording electrodes.

Methods: EEGs were recorded from 45 infants admitted to the neonatal intensive care unit (NICU). Three experts annotated seizures in EEG montages derived from 19, 8 and 4 electrodes. Differences between annotations were assessed by comparing intra-montage with inter-montage agreement (K).

Results: Three experts annotated 4464 seizures across all infants and montages. The inter-expert agreement was not significantly altered by the number of electrodes in the montage ($p = 0.685$, $n = 43$). Reducing the number of EEG electrodes altered the seizure annotation for all experts. Agreement between the 19-electrode montage ($K_{19,19} = 0.832$) was significantly higher than the agreement between 19 and 8-electrode montages ($dK = 0.114$; $p < 0.001$, $n = 42$) or 19 and 4-electrode montages ($dK = 0.113$, $p < 0.001$, $n = 43$). Seizure burden and number were significantly underestimated by the 4 and 8-electrode montage ($p < 0.001$). No significant difference in agreement was found between 8 and 4-electrode montages ($dK = 0.002$; $p = 0.07$, $n = 42$).

Conclusions: Reducing the number of EEG electrodes from 19 electrodes resulted in slight but significant changes in seizure detection.

Significance: Four-electrode montages for routine EEG monitoring are comparable to eight electrodes for seizure detection in the NICU.

© 2017 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Optimising neuro-critical care in the newborn has become a high priority issue in paediatrics (Abend et al., 2011a,b; Bonifacio et al., 2011; Glass and Rowitch, 2016). One of the key challenges in neuro-critical care is to provide brain monitoring that is continuous and uninterrupted in settings where personnel are changing constantly, neurophysiological support is limited, and an array of other medical devices are competing for cotside attention and

placements. These practical considerations have led neonatologists to use a two-electrode (single channel biparietal, P3–P4), or more recently, a four-electrode (two channel centro/fronto-parietal) derivation, to provide a measurement of neurological function (van Rooij et al., 2010a; Srinivasakumar et al., 2015). Such an approach has been questioned by clinical neurophysiologists because of their tradition of recording newborns with eight to ten electrodes, often referred to as “the full montage” (Hrachovy and Mizrahi, 2015; Shellhaas et al., 2011). There is, however, a limited amount of evidence for, or against, the idea that reducing the number of electrodes from a “full montage” would compromise performance in clinical settings (Shellhaas and Clancy, 2007; Wusthoff et al., 2009).

* Corresponding author at: BABA Center, Lastenlinna, P.O. Box 280, 00029 HUS, Helsinki, Finland.

E-mail address: nathan.stevenson@helsinki.fi (N.J. Stevenson).

The detection of seizures is perhaps the most widely used application of neonatal EEG monitoring (Boylan et al., 2013). In this case, the ideal recording configuration should strike a balance between clinical practicality and the fidelity of seizure detection. The comparison of different recording configurations for seizure detection is, however, not trivial as the presence of seizures cannot be absolutely defined. The present gold standard, visual EEG interpretation by an EEG expert, is inherently subjective and is further complicated by the assumption of perfect correlation between pathophysiological seizures and EEG grapho-elements. Subjectivity in visual interpretation will result in variability both within and between experts; hence, comparisons between recording configurations should be based on measuring intra- and inter-expert agreement.

The present study aimed to determine the effects of reducing the number of EEG electrodes on the visual identification of neonatal seizures. To do this, we have compared visual seizure identification of three human experts from a 19-electrode EEG montage with a conventional 8-electrode EEG montage and a 4-electrode EEG montage (van Rooij et al., 2010b; Srinivasakumar et al., 2015). We hypothesized that decreasing the number of electrodes would lead to a systematic and significant decrease in seizure detection.

2. Methods

2.1. Patients

EEGs were collected from neonates admitted to neonatal intensive care unit (NICU) at the Helsinki University Central Hospital, due to clinically suspected seizures between 2010 and 2014. The recordings were managed by the Department of Clinical Neurophysiology. The EEG signals were collected using NicOne EEG amplifier (sampling frequency of 256 Hz; Cardinal Healthcare/Natus, USA) from EEG caps (sintered Ag/AgCl electrodes; Waveguard, ANT-Neuro, Germany) with 19 electrodes positioned as per the international 10–20 standard (further details of the neonatal EEG recording method, see Vanhatalo et al. (2008) and <http://www.babacenter.fi/en/methods/eeeg/>). The initial survey of EEG reports identified 87 neonates where a 19-electrode EEG recording was requested due to clinical suspicion of seizures. Preliminary inspection of these EEGs found 45 neonates with evidence of seizure. The demographics of this cohort are shown in Table 1. The total duration of EEG annotated was 3962 min. The study was approved by the Ethics Committee of the Helsinki University Children's Hospital. All patient identifying information was removed from each recording before annotation.

Table 1

Demographics of the cohort. Summary measures are median (interquartile range) unless otherwise stated ($n = 45$).

Gestational age (weeks)	39.1 (38.0–40.6)
Birthweight (g)	3451 (2830–3836)
Gender (M/F)	24/21
Postmenstrual age at EEG (weeks)	40.0 (39.0–41.4)
Postnatal age at EEG (days)	2.0 (1.0–5.5)
EEG recording duration (min)	77.1 (63.1–100.0)
Diagnosis	
HIE/Asphyxia	25
HIE/Asphyxia and Stroke	2
Stroke	6
Other	8
Unknown	4

HIE – hypoxic ischaemic encephalopathy.

EEG – electroencephalogram.

Other includes infections, brain anomalies and kernicterus.

2.2. Pre-processing and seizure annotation

Four copies of each EEG recording were created with bipolar montages derived from the full 19 electrode recording (see Fig. 1). This resulted in EEG recordings with 18 (two copies), 8 (one copy), and 4 (one copy) derivations, respectively (see Table S1 in the Supplementary Material). The EEG file names were then randomised, with different randomisations for each of the four EEG copies to ensure a fully blinded review.

Each EEG file was annotated for seizures by three experts, each with over 10 years of experience in the visual interpretation of neonatal EEG (LL, JM, SV). Each expert was requested to accurately annotate seizures identified using the well-established definition: a distinct, abnormal electrographic event with a clear beginning and end comprising sustained, repetitive evolving spike/sharp waves or rhythmic waveforms (Clancy et al., 1988). Annotations included the start and end of each seizure considering all available channels. Initial settings were a paper speed of 30 mm/s, a sensitivity of 100 μ V/cm with frequency cutoffs of 0.5 Hz (low) and 70 Hz (high). Reviewers were permitted to alter these settings during review.

Each expert was blinded to the clinical details of infants, and was only aware that the neonate had suspected EEG seizures and were recorded two to six years earlier. Video, additional polygraphic channels, apart from the electrocardiogram, and technician's comments were not available during review. The first round of annotations included the 19-electrode montage, the 8-electrode montage and the 4-electrode montage. In order to reduce the confounding effect of previous exposure to EEG recordings, the second copy of the 19-electrode montage was annotated at least 3 months after the completion of the first round of annotation.

2.3. Analysis

2.3.1. Primary analysis

The main objective was to determine the effect on the visual identification of seizures from reducing the number of recording electrodes. We assumed that intra-montage agreement within the 19-electrode montage yields a benchmark measure of subjectivity. To this end, we compared intra-montage agreement (19-electrode) with inter-montage agreement (19 vs 8 electrodes and 19 vs 4 electrodes) to determine differences attributable to a reduction in the number of recording electrodes. We also compared the inter-expert agreement of the 19-electrode annotation to the inter-expert-montage agreement (19 vs 8 electrodes and 19 vs 4 electrodes) to determine if any differences were in excess of inter-expert agreement.

2.3.2. Secondary analysis

The effect of reducing the number of electrode in the EEG montage was further quantified using the agreement rate of three measures that could be estimated from the annotations of the experts: (1) seizure diagnosis, (2) seizure number, (3) seizure burden. Agreement in seizure diagnosis was defined if any seizure was annotated in the recording of an infant, irrespective of temporal alignment, across experts. Agreement in seizure number was defined when seizure events had any overlap across experts. Agreement in seizure burden was defined with a second by second comparison of annotations across experts. The visual inspection of disagreements between montages was also performed after these analyses to determine common factors in disagreements.

2.3.3. Statistical analyses

We used Cohen's kappa (K) statistic to measure the inter- and intra- montage agreement in the annotations per expert (Cohen, 1968). Inter-expert agreements were estimated across three

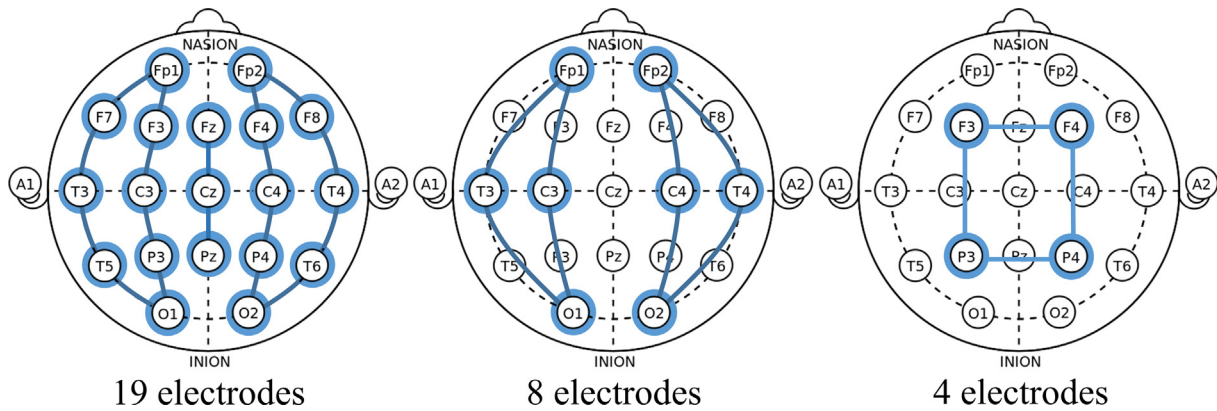


Fig. 1. The electrode placements used in this study. The 19 electrodes, 8 electrodes and 4 electrodes were used to generate bipolar montages of 18, 8, and 4 channels, respectively, for annotation by the human expert.

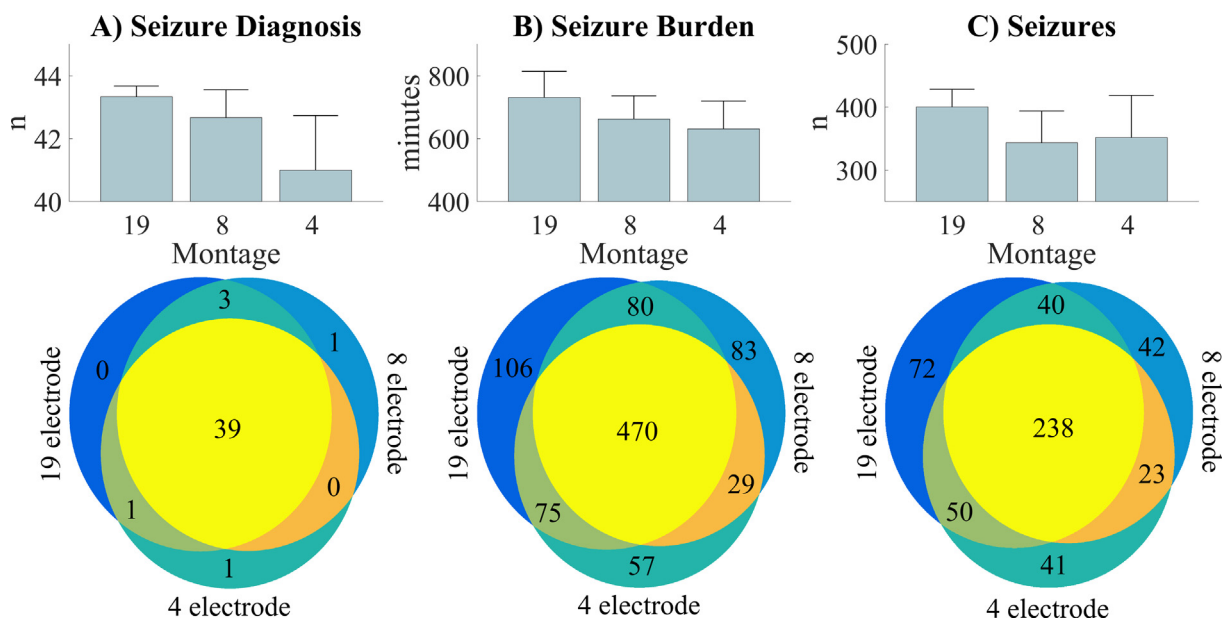


Fig. 2. Inter-montage agreement between 19, 8 and 4 electrode montages for interpreting the EEG with respect to seizure diagnosis (A), seizure burden (B), and seizures (C). Results were averaged across the 3 experts. Venn diagrams summarize the full comparison of montages. Note that there are individuals and seizures, as well as seizure burden, that do not overlap between 4 and 8 electrode recordings. Error bars reflect the standard error on the mean estimated across experts. SB – seizure burden.

experts using Fleiss' kappa (K) statistic (Fleiss, 1971). Kappa was calculated on the raw annotations where agreements were evaluated per second and then averaged across the full EEG recording. Kappa values were estimated on individual infants and on a single annotation of each infant linked together (concatenated). The latter was used to ensure all infants were included in analysis as individual estimates could not be performed on infants where at least one expert could not identify seizures in the EEG recording. Values were, therefore, summarised over individuals with a median and interquartile range or over the concatenated annotation with the value and a 95% confidence interval estimated with a bootstrap.

Differences in agreement between montages were assessed with Hotelling's T^2 statistic under the null hypothesis that the difference between intra-montage and inter-montage agreement was zero (Hotelling, 1951). Differences between the inter-expert agreement over montages were tested with a one-way ANOVA under the null hypothesis of no difference in kappa with respect to montage. Post hoc analysis was performed and is included in the

Supplementary Material. Infants where kappa could not be calculated were excluded from these analyses. Variables that were heavy tailed were transformed using the inverse hyperbolic sign function (Burbidge et al., 1988). Two-sided tests were used and a p -value less than 0.05 was deemed significant. The association between agreement and seizure burden (log-transformed) was evaluated with a linear correlation coefficient.

3. Results

3.1. Agreement between experts

The three experts annotated a total of 4464 seizures across infants and montages. The agreement between experts was not significantly different when interpreting EEG recordings with different electrode numbers ($p = 0.685$, $n = 43$). The inter-expert agreement was 0.745 (95%CI: 0.645–0.819), 0.745 (95%CI: 0.644–0.825), and 0.731 (95%CI: 0.605–0.822) for the 19, 8, and

Table 2
Agreement between the visual interpretations of experts for different EEG montages. Agreement is measured with Cohen's kappa value calculated on the concatenated recording. 19 vs 19 denotes intra-montage agreement for the 19-electrode montage and the remaining rows denote inter-montage agreement. The 95% confidence interval is in parentheses and is estimated using a bootstrap.

Electrode comparison	Expert A	Expert B	Expert C
19 vs 19	0.863 (0.803–0.905)	0.809 (0.738–0.870)	0.824 (0.734–0.888)
19 vs 8	0.792 (0.711–0.846)	0.727 (0.653–0.790)	0.722 (0.605–0.813)
19 vs 4	0.771 (0.687–0.835)	0.754 (0.681–0.816)	0.752 (0.664–0.816)
8 vs 4	0.786 (0.708–0.838)	0.710 (0.635–0.772)	0.694 (0.595–0.768)

Table 3
The agreement in measurements based on the annotation of the 4 and 8-electrode montages compared to the 19-electrode montage as 'gold standard'. The 19-electrode column defines intra-montage agreement. False detections were defined as number per hour. Results are presented as median (interquartile range) across the cohort and values are averaged across the three experts. For measures of seizure burden and seizure number only infants with a consensus (across experts) diagnosis of seizure were included.

	19-electrode	8-electrode	4-electrode
Diagnosis (<i>n</i> = 45)	97%	96%	89%
Seizure burden (<i>n</i> = 41)	91% (76–98%)	90% (72–96%)	88% (73–96%)
Seizure number (<i>n</i> = 41)	91% (82–100%)	84% (57–100%)	79% (53–98%)
False detections (<i>n</i> = 45)	0.59 (0.12–1.26)	0.43 (0.0–0.93)	0.31 (0.13–0.86)

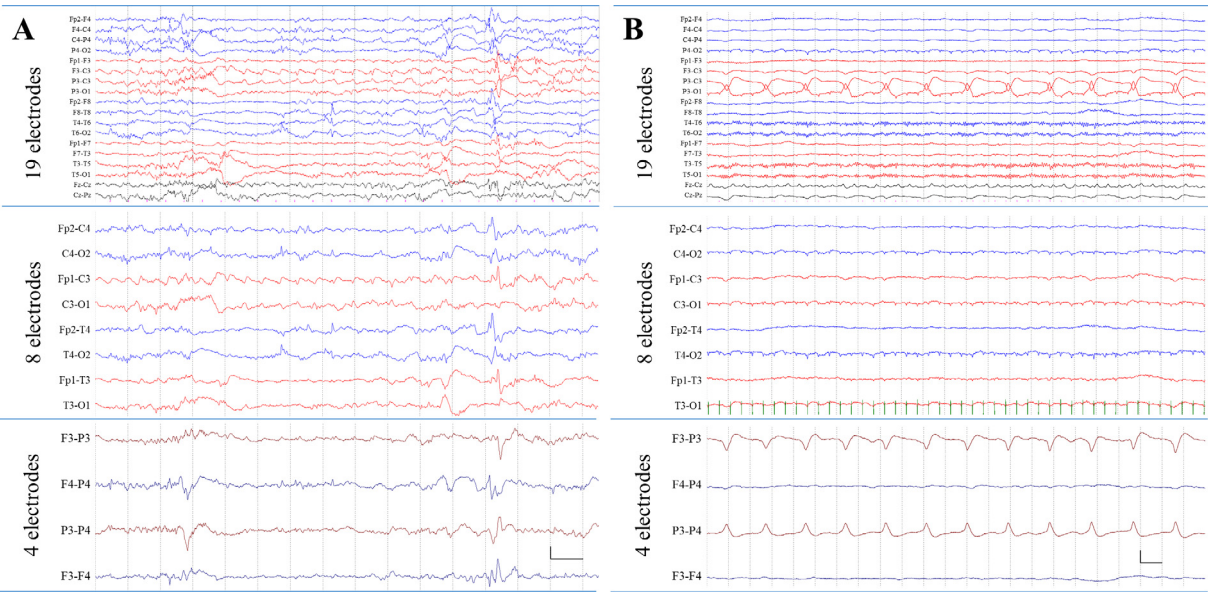


Fig. 3. EEG patterns resulting in reduced intra-montage agreement. (A) A focal seizure at central area in an infant with otherwise normal amplitude background activity. The seizure is clear in the 8-electrode recording, and is less obvious in the 4-electrode recording. (B) A focal seizure at parietal area in an infant with otherwise abnormally low amplitude background activity. The seizure is clear in the 4-electrode recording, but there are no manifestations in the 8-electrode recording as the parietal electrode is missing. The calibration (1 s, 100 μ V) is identical within, but different between, (A) and (B).

4 electrode montages, respectively. A summary of the seizure annotations for each expert and montage, the agreement between each annotation, and the bias and prevalence indices of comparisons are given in [Tables S2, S3 and S6 of the Supplementary Material](#).

3.2. Agreement between EEG montages

There was substantial agreement between the annotations of EEG montages with a different number of recording electrodes ([Fig. 2](#)). The average agreement across experts between the annotation of 19 and 8, 19 and 4, and 8 and 4 electrode montages was 0.747 (95%CI: 0.669–0.814), 0.759 (95%CI: 0.682–0.819), and 0.730 (95%CI: 0.659–0.788) respectively. Reducing the number of electrodes in the EEG montage, nevertheless, resulted in a significantly different

annotation of seizures. Compared to the intra-montage agreement of the 19-electrode montage ($K = 0.832$, 95%CI: 0.776–0.878), the inter-montage agreement between 19 and 8, and 19 and 4-electrode montages were significantly lower (8-electrode, $dK = 0.114$; $p < 0.001$, $n = 42$ and 4-electrode, $dK = 0.113$; $p < 0.001$, $n = 43$; [Table 2](#)). This reduction in agreement was, however, not significantly different between 8 and 4-electrode montages ($dK = -0.002$; $p = 0.07$, $n = 42$; [Table 2](#)). There was also a significantly lower seizure burden annotated in the 8- and 4-electrode montage compared to the 19-electrode montage (8-electrode $dSB = 1.9$ min, $p < 0.001$, $n = 45$; 4-electrode $dSB = 2.5$ min, $p < 0.001$, $n = 45$). The reduction of seizure burden was, however, not significantly different between 8 and 4-electrode montages ($dSB = -0.7$ min, $p = 0.20$, $n = 45$). There was a significant reduction in the number of seizures annotated in the 4- and 8-electrode

Table 4

Seizure characteristics for infants with low and high inter-expert agreement. Temporal characteristics were averaged across all potential annotations per infant and summarised with the median (interquartile range).

	High agreement (n = 25)	Low agreement (n = 20)
Seizure burden (min)	14.9 (7.5–34.2)*	5.4 (2.5–10.9)
Seizure number	5.4 (2.6–11.8)	5.3 (2.4–12.4)

* Significantly higher ($p < 0.05$; Mann-Whitney U test).

montage compared to the 19-electrode montage (8-electrode $dN = 1.2$, $p < 0.001$, $n = 45$; 4-electrode $dN = 1.0$, $p < 0.001$, $n = 45$). The reduction of seizure number was, however, not significantly different between 8 and 4-electrode montages ($dN = -0.2$, $p = 0.166$, $n = 45$). Post hoc analysis of the results in this section are shown in [Tables S4 and S5 of the Supplementary Material](#).

The agreement between the annotation of the 19-electrode montage and 4- and 8-electrode montage was also significantly lower than the agreement between expert annotation of the 19-electrode montage (8-electrode $dK = 0.056$, $p < 0.01$, $n = 42$; 4-electrode $dK = 0.075$, $p < 0.001$, $n = 42$).

3.3. Agreement with the 19-electrode montage

Out of 400 seizures identified in the 19-electrode montage, 70% were detected by the 8-electrode montage, 72% were detected by the 4-electrode montage, and 60% were detected by both ([Fig. 2C](#)). 18% of seizures detected in the 19-electrode montage were not detected in the 8- or 4-electrode montage. Out of the 343 seizures detected in the 8-electrode, 76% were also detected in the 4-electrode montage and 12% were not detected in either the 4 or 19 electrode montage ([Fig. 2C](#)). Summary values across infants are shown in [Table 3](#).

3.4. Analysis of disagreements

We carried out a visual *post hoc* EEG review to identify possible reasons for disagreements between experts or montages in some infants. We could readily identify three different reasons: (i) Focal seizures in the occipital, temporal or parietal lobes may go unnoticed in the reduced electrode montages; (ii) Seizures with a low amplitude on the EEG relative to background activity; (iii) Seizures that may resemble slow rhythmic artefacts requiring multiple channels to confirm. There were 20 infants where inter-montage agreement was less than 0.5 for at least two out of 12 potential montage comparisons. In half of these infants ($n = 10$), seizures were focal, seven had low amplitude seizures with respect to background and three had EEG seizures that resembled slow rhythmic artefacts when assessed from one channel ([Fig. 3](#)). Temporal characteristics of seizures in the low and high agreement groups are shown in [Table 4](#). Infants with low agreement had a significantly lower annotated seizure burden ($p < 0.05$). Seizure burden was also correlated with agreement in 5/12 comparisons; the median correlation was 0.271 (IQR: 0.147–0.358; $n = 12$, see [Fig. S1 in the Supplementary Material](#) for an example).

4. Discussion

Our results show that visual interpretation of 8 and 4-electrode EEG montages for seizure detection has substantial agreement with the interpretation of a 19-electrode EEG montage. The annotation of reduced electrode montages was, nevertheless, significantly different from the 19-electrode montage resulting in a reduced estimate of seizure burden. These differences were significantly higher than inter-expert disagreement. The losses in

seizure detection from 19-electrode recordings were no different between the conventional 8-electrode neonatal recording and the 4-electrode recording now widely used in the NICU.

To our knowledge, this is the first study that quantifies the differences in visual seizure detections by benchmarking with a combination of intra- and inter-expert agreements. This approach controls for inter-expert agreement which means that the reduction in seizure detection is real and cannot be explained by the subjectivity of visual interpretation. Our findings generally support prior studies that even a limited channel recording with central or parietal electrodes identifies the majority of seizures ([Shellhaas and Clancy, 2007](#); [Bourez-Swart et al., 2009](#)). Our present work extends prior knowledge by showing (i) that the current state of art in neonatal EEG, an 8-electrode recording, is already compromised compared to the 19-electrode recording that is routinely used in older patients, and (ii) that there is no apparent gain in seizure detection by increasing the number of recording electrodes from four to eight.

Our findings indicate that decreasing electrode numbers result in reduced estimates in seizure burden. As inter-expert agreement did not significantly change with the number of recording electrodes, the reduction in seizure burden must be due to the reduced manifestation of seizures with limited spatial extent on the decreased electrode montage. Our prior theoretical and empirical work found that the neonatal EEG is spatially rich with distinct electrical activity recorded at electrodes placed only few centimetres apart ([Odabae et al., 2013](#)). The reduction in seizure burden with the number of recording electrodes was, however relatively small, and was no different between 4 and 8 electrode montages. This suggests that while some seizures may be focal, the majority of seizures spread wider on the cortical surface ([Shellhaas and Clancy, 2007](#); [Wusthoff et al., 2009](#); [Nagarajan et al., 2011](#)). Indeed, this idea was confirmed in our post hoc analysis of infants with low agreement; seizures tend to be missed if they were focal in an area not covered by the reduced number of electrodes such as parietal areas for the 8-electrode montage and occipital areas for the 4-electrode montage. This spatial patterning of EEG seizures implies that there may be an EEG montage that minimises the number of electrodes and maximises seizure detection.

The present study has some inherent limitations. We selected EEG recordings from infants with suspected EEG seizures which may lower the threshold for visual seizure detection. We have also not included infants who were not suspected of having seizure. This group of infants may be of use when studying inter-expert agreement but are not necessarily required when evaluating inter-montage agreement which is the primary aim of this study. The use of short term recordings may place downward pressure on measures of inter-expert agreement due to limited total seizure burden in many infants. Low seizure burden is associated with decreased inter-expert agreement and may explain why the values of inter-expert agreement reported in our study are at the lower bounds of those reported in the literature ([Smit et al., 2004](#); [Shah et al., 2008](#); [Stevenson et al., 2015](#)). The use of short duration recordings provided reduced opportunity for all experts to annotate seizures in a recording. As a result, for certain analyses infants had to be excluded as kappa values could not be estimated (the calculation of the hypothetical probability of chance agreement component of kappa was not possible). This potential source of bias does not affect our results based on agreement as the analysis of concatenated recordings of all infants (a method that overcomes the issue of infant exclusion – see the bootstrap analysis in [Tables S4 and S5 in the Supplementary Material](#)) supports our conclusions which suggests that infants were excluded not because of a trend towards low agreement but rather poor seizure prevalence. There are several possible permutations for reduced electrode montages and it is possible that the specific montages used were not optimal.

The fact that our reduced channel montages captured a high proportion of seizure events, however, suggest that the montages used were sufficient. EEG interpretation was performed on the multi-channel EEG with an additional electrocardiogram channel. This is not typical as EEG is commonly interpreted in the context of notes from the technician, polygraphy and video. These issues may slightly bias the numerical estimates of agreement; however, they should not significantly confound the relative differences between measures used in this study.

There is, currently, no method that can conclusively confirm the presence or absence of neonatal seizures. The best available benchmark is visual EEG interpretation by the human expert. It is now well established that the inter-expert agreement is not perfect for the detection of a range of EEG patterns (Scher et al., 1994; Abend et al., 2011a,b; Stevenson et al., 2015; Wusthoff et al., 2017). In an era of precision medicine, the important clinical question is how tarnished can the supposed gold standard become and still be useful for neuro-critical care, or be used as an outcome measure in clinical trials (Halford et al., 2016). We show that reducing the number of electrodes in an EEG montage from 19 to 8 or 4 significantly changes the annotation of seizures beyond inter-expert differences, although the relationship between seizure detection and electrode number is nonlinear. While the reduction in seizure detection (both in terms of seizures and burden) is relatively minor and potentially inconsequential in day-to-day practice, it may be important when a higher level of accuracy is required such as when evaluating outcome measures of neonatal seizures in clinical trials. The fact that there was no significant difference in seizure detection when comparing the 8-electrode and 4-electrode recordings to the 19-electrode recordings is clinically reassuring and compatible with the idea that long-term EEG monitoring with the now widely adopted 4-electrode routine has validity. This is important for clinical environments where the resources available for EEG monitoring are limited.

Acknowledgements

This study was supported by the Finnish Cultural Foundation, EU Marie Skłodowska Action (H2020-MCSA-IF-656131), Academy of Finland (#276523 and #288220), and the Sigrid Juselius Foundation. The authors also wish to acknowledge the support of Mr. Jarmo Maki for annotating the EEG recordings and Mr. Panu Oksanen for collecting the dataset of EEG recordings from clinical archives.

Conflict of interest

None of the authors have potential conflicts of interest to be disclosed.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.clinph.2017.10.031>.

References

- Abend NS, Gutierrez-Colina A, Zhao H, Guo R, Marsh E, Clancy RR, et al. Interobserver reproducibility of electroencephalogram interpretation in critically ill children. *J Clin Neurophysiol* 2011a;28:15–9.
- Abend NS, Topjian AA, Gutierrez-Colina AM, Donnelly M, Clancy RR, Dlugos DJ. Impact of continuous EEG monitoring on clinical management in critically ill children. *Neurocritical Care* 2011b;15:70–5.
- Bonifacio SL, Glass HC, Ploquin S, Ferriero DM. A new neurological focus in neonatal intensive care. *Nat Rev Neurol* 2011;7:485–94.
- Bourez-Swart MD, van Rooij L, Rizzo C, de Vries LS, Toet MC, Gebbink TA, et al. Detection of subclinical electroencephalographic seizure patterns with multichannel amplitude-integrated EEG in full-term neonates. *Clin Neurophysiol* 2009;120:1916–22.
- Boylan GB, Stevenson NJ, Vanhatalo S. Monitoring neonatal seizures. *Sem Fetal Neonatal Med* 2013;18:202–8.
- Burbidge JB, Magee L, Robb AL. Alternative transformations to handle extreme values of the dependent variable. *J Am Stat Assoc* 1988;83:123–7.
- Clancy RR, Legido A, Lewis D. Occult neonatal seizures. *Epilepsia* 1988;29:256–61.
- Cohen J. Weighed kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychologic Bull* 1968;70:213–20.
- Fleiss JL. Measuring nominal scale agreement among many experts. *Psychologic Bull* 1971;76:378–82.
- Glass HC, Rowitch DH. The role of the neurointensive care nursery for neonatal encephalopathy. *Clin Perinatol* 2016;43:547–57.
- Halford JJ, Arain A, Kalamangalam GP, LaRoche SM, Bonilha L, Basha M, et al. Characteristics of EEG interpreters associated with higher inter-rater agreement. *J Clin Neurophysiol* 2016. <https://doi.org/10.1097/WNP.0000000000000344>.
- Hottelling H. A generalized t test and measure of multivariate dispersion. In: Second Berkeley symposium on mathematical statistics and probability, vol. 1; 1951. p. 23–41.
- Hrachovy RA, Mizrahi EM. Atlas of neonatal electroencephalography. 4th ed. New York: Demos Medical; 2015.
- Nagarajan L, Ghosh S, Palumbo L. Ictal electroencephalograms in neonatal seizures: characteristics and associations. *Ped Neurol* 2011;45:11–6.
- Odabae M, Freeman WJ, Colditz PB, Ramon C, Vanhatalo S. Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes. *Neuroimage* 2013;68:229–35.
- Scher MS, Bao QH, Mazumdar S, Painter MJ, Alvin J, Redmond CK. A reliability study of a neonatal seizure scoring system. *J Epilepsy* 1994;7:273–8.
- Shah DK, Mackay MT, Lavery S, Watson S, Harvey AS, Zempel J, et al. Accuracy of bedside electroencephalographic monitoring in comparison with simultaneous continuous conventional electroencephalography for seizure detection in term infants. *Pediatrics* 2008;121:1146–54.
- Shellhaas RA, Clancy RR. Characterization of neonatal seizures by conventional EEG and single-channel EEG. *Clin Neurophysiol* 2007;118:2156–61.
- Shellhaas RA, Chang T, Tsuchida T, Scher MS, Riviello JJ, Abend NS, et al. The American Clinical Neurophysiology Society's guideline on continuous electroencephalography monitoring in neonates. *J Clin Neurophysiol* 2011;28:611–7.
- Smit LS, Vermeulen RJ, Fetter WP, Strijers RL, Stam CJ. Neonatal seizure monitoring using non-linear EEG analysis. *Neuropediatrics* 2004;35:329–35.
- Srinivasakumar P, Zempel J, Trivedi S, Wallendorf M, Rao R, Smith B, et al. Treating EEG seizures in hypoxic ischemic encephalopathy: a randomized controlled trial. *Pediatrics* 2015;136:e1302–9.
- Stevenson NJ, Clancy RR, Vanhatalo S, Rosén I, Rennie JM, Boylan GB. Interobserver agreement for neonatal seizure detection using multichannel EEG. *Ann Clin Translat Neurol* 2015;2:1002–11.
- van Rooij LG, Toet MC, van Huffelen AC, Groenendaal F, Laan W, Zecic A, et al. Effect of treatment of subclinical neonatal seizures detected with aEEG: randomized, controlled trial. *Pediatrics* 2010a;125:e358–66.
- van Rooij LG, de Vries LS, van Huffelen AC, Toet MC. Additional value of two-channel amplitude integrated EEG recording in full-term infants with unilateral brain injury. *Arch Dis Child-Fetal Neonatal Ed* 2010b;95:F160–8.
- Vanhatalo S, Metsäranta M, Andersson S. High-fidelity recording of brain activity in the extremely preterm babies: feasibility study in the incubator. *Clin Neurophysiol* 2008;119:439–45.
- Wusthoff CJ, Shellhaas RA, Clancy RR. Limitations of single-channel EEG on the forehead for neonatal seizure detection. *J Perinatol* 2009;29:237–42.
- Wusthoff CJ, Sullivan J, Glass HC, Shellhaas RA, Abend NS, Chang T, et al. Interrater agreement in the interpretation of neonatal electroencephalography in hypoxic-ischemic encephalopathy. *Epilepsia* 2017;58:429–35.